The Marriage Matching Problem with Information Limited By Social Networks

Debra Hevenstone
University of Bern

Nov 30, 2013

1 Introduction

The shape of sexual networks (who has slept with whom) is of key interest to public health officials and epidemiologists. Networks connected by a few central actors suggest targeted interventions (e.g. HIV testing at clubs), while broadly connected networks suggest blanket approaches. Because network data analysis requires high response rates and clear community boundaries, very little empirical research has been done, and what has been done is in environments with clear boundaries, like schools (Bearman et al., 2004). At least among youth in the National Longitudinal Study of Adolescent Health data (Add Health), sexual history networks take the shape of a "spanning tree," i.e. many people are connected through a single large component. Bearman et al. (2004) suggests that the mechanism generating this sexual network is random mixing with homophilous preferences and a taboo on dating one's ex's ex's ex. However, there is evidence suggesting that a different mechanism might be at work: romantic relationships form over social networks. If that is the case, the spanning tree may be specific to youth who form relationships in environments with closed boundaries and tightly connected friendship networks.

In a first step, this paper uses simulation to generate sexual history networks. The first simulations allow individuals to match with all potential partners in the dating pool, taking homophily into account and varying the amount of time individuals can date. In a second step "limited scope" search is tested, limiting individuals' dating pool to their 2-hop social network. The resulting simulated sexual networks are then compared to the Add Health results. We hypothesize that when search scope and search length are limited, a spanning tree will be observed. In contrast, if dating is allowed to continue in the limited scope environment, loops will emerge, and if full scope search is allowed (particularly in a larger population), no main connected component will emerge.

In a second step, this paper uses social network data to test whether romantic relationships form over social networks. A recent survey collected in a student exchange program measured the progressive establishment of relationships, including romantic relationships. This offers a unique opportunity to examine how social and romantic networks co-evolve, given that at the start of the program there were no existing relationships. As the data was collected after only five months of interaction, it is unlikely that the dating history network will be a spanning tree, as observed in the Add Health data which spanned 18 months. However, the dynamic data allows us to estimate the probability of a romantic edge forming, given potential partners' distance across the network (i.e. is dating scope limited by one's social network?).

2 Background

Research on sexual networks in an American high school suggest that sexual networks are characterized by long chains or "spanning trees", meaning that a large part of the school had had sexual contact with each another (Bearman et al., 2004). One mechanism shown to be capable of generating a sexual network similar in terms of density at maximum reach, size of the largest component, centralization, mean geodesic length, and the number of four cycles, is random mixing with homophily and a taboo on dating one's ex's ex's ex.

While there is evidence that people mix based on homophily, there is empirical evidence that people do not mix randomly. A comparison of on-line and off-line dating has shown that on-line, where people have a large pool to select from, they tend to match with less homophily on certain traits, like education (Gunter et al., 2010). One potential reason for the difference between on and off line dating is that people meet potential partners through their social networks, which are characterized by strong homophily.

Empirically it is extremely difficult to determine how social networks impact relationship formation. At the time we observe a couple and their social network, we generally do not know whether they met through their mutual friends or whether they have mutual friends because they are in a relationship. It has been shown that one can identify a relationship edge in a person's ego network based on "dispersion" or the number of dyads in an ego network that no longer have a mutual friend once the ego and suspected partner are removed from the network (i.e. we can guess who the ego's partner is because their partner bridges their various social groups) (Backstrom and Kleinberg, 2014). If we could observe an ego network at the time when the ego meets his or her partner, we might guess that embeddedness the number of friends the potential couple has in common) might be a better predictor of a relationship.

In this paper we examine whether relationships form over social networks. Meeting partners over a social network should yield a dating history network that resembles a spanning tree only if matching occurs in a limited time frame and in a closed environment with a clear boundary (i.e. high school). Among adults in a boundary-less environment with longer-term matching, dating over one's social network should yield a different sexual network.

3 Simulation

The stable marriage problem (SMP) is the problem of finding a stable matching between two sets of elements given a set of preferences for each element. It is conceived of as the "marriage problem" because we might imagine a world with men who have preferences in women, and women who have preferences in men, and we want them all to pair off. Formally: Given n men and n women, where each person has ranked all members of the opposite sex with a unique number between 1 and n in order of preference, marry the men and women together such that there are no two people of opposite sex who would both rather have each other than their current partners. If there are no such people, all the marriages are "stable" or in other words, no 2 people would both be happier with each other than with their current partner.

One algorithm that can be used to find a stable solution is the "Gale-Shapely Algorithm." In this algorithm there are multiple rounds in which men can propose to their most-preferred woman to whom they have not yet proposed. Women accept a suitor if he improves her current utility, but if another, better, suitor comes around, the woman may reject her suitor for the new one. Within $n2 - n + 1$ rounds a stable solution is found (where n is the number of men or women).

We simulate a pool of men and women with four randomly assigned characteristics per agent: two characteristics (we can call them "attractiveness" a_i and "intelligence" s_i) from a normal distribution, $\mu = 5$, $\sigma = 1$, preference for partner's attractiveness (from a uniform distribution $(0, 1)$) α_i, and preference for partner's intelligence ($1 - \alpha_i$). Men and women rank the members of the opposite sex, calculating their potential utility using a "Cobb Douglas" function with the coefficients constrained to sum to 1, meaning there are constant returns to scale. Utility for person i considering dating person j is:

$$u_i = a_j^{\alpha_i} * s_j^{1-\alpha_i} \tag{1}$$

In the real world, people do not have the chance to make $n2 - n + 1$ offers. For this reason we constrain dating to various rounds of matching. The resulting sexual history networks are then compared to the Add Health data using the same descriptive measures used in (Bearman et al., 2004): density at maximum reach, size of the largest component, centralization, mean geodesic length, and the number of four cycles.

In the real world people do not only have limited time, they also have limited scope. As such, we re-run the same same experiment, but this time, men are only allowed to make offers to women within 2 steps in their social network. Before running this second experiment we need to generate a social network. Empirically social networks are characterized by homophily (meaning that people have friends like themselves), by triad closure (people tend to make friends with friends of friends), and by preferential attachment (people are likely to make friends with those who already have many friends). One can empirically measure these factors using either an exponential graph model or an actor based model Snijders (2001) with the predicted probability that 2 people might be friends as follows:

$$Pr_{i \cdot j=1} = \frac{e^f}{1 + e^f} f = \boldsymbol{\beta}(\mathbf{x_i} + \mathbf{x_j}) + \boldsymbol{\gamma} x_{ij} + \psi(D_i + D_j) + \gamma T_{ij} \tag{2}$$

Where x_i are individual characteristics, x_{ij} are similarity of characteristics, D_i is current degree, T_{ij} is number of friends in common, and $\boldsymbol{\beta}, \boldsymbol{\gamma}, \psi$, and γ are coefficients estimating the strength of each effect. These coefficients can be estimated using social network data and then used to simulate the social networks. Currently the simulation assumes the average degree is 4, the coefficients on attractiveness and intelligence are .03, the coefficient for popularity effects is .1, the coefficients on the dyad's difference in attractiveness and intelligence is -.1, and the coefficient on the number of friends in common is .1.

In sum, a Gale Shapely matching algorithm is run with full and limited scope searches using a range of time limits (the number of offers the men can make).

4 Empirical Analysis

Students at the University of Bern are currently collecting network data on exchange students, including the type of relationship (friend, colleague, roommate, romantic partner) and the month in which the relationship was established.

As a starting point this data will be used to estimate a mixed effects logistic model, in which each observation is a cross-gender dyad of individuals i and j in a given month, t with crossed effects by each of the individuals who are a member of the dyad.

$$ln(\frac{\pi_t}{1 - pi_t}) = \beta_0 + \gamma X_{i,(t-1)} + \delta X_{j,(t-1)} + \tau X_{ij,(t-1)} + \nu F_{ij,(t-1)} + \alpha \quad (3)$$

X_{it} is a vector of individual characteristics for individual i in time period t, $X_{ij,(t-1)}$ are dyadic characteristics in period $t-1$, and $F_{ij,(t-1)}$ are the number of friends in common between i and j in period $t-1$. The coefficient, ν will suggest whether having friends in common was a significant predictor of establishing a romantic relationship.

Future work will use traditional network methods to measure the probability of romantic edges, including an exponential random graph model and an actor-based model. Notably, the existing "canned" models are not designed to look at the co-evolution of friendship and dating networks. The simulation developed in part 1 could, and should, be extended. It is already an actor-based model of the co-evolution of friendship and dating. However, the current simulation can not yet be used to fit empirical data, as it is not yet programmed in such a way as to adjust the coefficients using MCMC, nor use the macro-descriptive network statistics as a test statistic to compare with empirical data.

In future work the friendship data can also be used to roughly estimate the coefficients used in generating the simulated friendship networks. The collected data does not include attractiveness nor intelligence or preferences but does include measures of individuals' study effort, the number of friends they have, the number of hook ups they have had, and the same traits among all their friends in the exchange program. As such, the data can be used to look at the comparative impact of individual traits like studiousness and promiscuity, dyadic similarity on the same traits, triad closure, and popularity. These coefficients can be estimated using existing "canned" actor-based models.

5 Results and Conclusion

We anticipate that the simulation results will suggest that only under the conditions of a limited time and scope search will dating histories take the form of a spanning tree. We also anticipate that the empirical data will show that dating networks form over friendship networks.

References

Backstrom, Lars and Jon Kleinberg. 2014. "Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook." Technical report, Proc. 17th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW).

Bearman, Peter S., James Moody, and Katherine Stovel. 2004. "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks." *American Journal of Sociology* 110:4491.

Gunter, J. Hitsch, Ali Hortasu, and Dan Ariely. 2010. "Matching and Sorting in Online Dating." *American Economic Review* 100:130–63.

Snijders, T. C. 2001. "The statistical evaluation of social network dynamics." *Sociological Methodology* 31:361–395.